

**Remarks**

Reconsideration of this Application is respectfully requested.

Upon entry of the foregoing amendment, claims 1-5, 8-13, 16-19, 22-26, and 29-50 are pending in the application, with Claims 1, 16, 29, 39, 49 and 50 being the independent claims. Claims 6-7 and 20-21 are sought to be cancelled without prejudice to or disclaimer of the subject matter therein. Claims 14-15 and 27-28 were previously cancelled. New claims 49-50 are sought to be added. These changes are believed to introduce no new matter, and their entry is respectfully requested.

Based on the above amendment and the following remarks, Applicants respectfully request that the Examiner reconsider all outstanding rejections and that they be withdrawn.

***Arguments in Previous Office Action Response and Examiner Interview***

Applicants once again extend their appreciation to the Examiner for meeting with Applicants' counsel on November 10, 2004. During that meeting, Applicants' counsel explained the invention with a particular focus on independent claims 1 and 29. In particular, Applicants' counsel explained how a segment-level actual usage value for one or more word combinations can be determined, and also explained how a segment-level expected usage value for one or more word combinations can be computed.

During the Interview, Applicants' counsel demonstrated that U.S. Patent No. 6,678,694, entitled *Indexed, Extensible, Interactive Document Retrieval System*, issued to Zimmerman, et al. on Jan 13, 2004 ("Zimmerman Patent") and U.S. Patent No. 5,745,776, entitled *Enhanced Electronic Dictionary*, issued to Charles Sheppard, II on

Apr. 28, 1998 ("Sheppard Patent") did not render the claims of the present application obvious under 35 U.S.C. § 103(a).

Additionally, Applicants provided an Amendment and Remarks, dated November 12, 2004 ("Nov. 12<sup>th</sup> Amendment and Remarks"), in response to the Office Action received from the Examiner, dated July 14, 2004 ("First Office Action"). In the Nov. 12<sup>th</sup> Amendment and Remarks, Applicants further demonstrated that the claims, as amended, should be allowable over the references cited by the Examiner in the First Office Action. Specifically, the Examiner had provided rejections under 35 U.S.C. § 103 based on various combinations of the Zimmerman Patent, the Sheppard Patent, U.S. Patent No. 6,446,061, entitled *Taxonomy Generation for Document Collections*, issued to Doerre et al., on Sep. 3, 2002 ("Doerre Patent") and U.S. Patent No. 6,038,560, entitled *Concept Knowledge Base Search and Retrieval System*, issued to Kelly Wical on Mar. 14, 2000 ("Wical Patent").

In the Present Office Action, the Examiner has introduced a new reference – U.S. Patent Application Publication No. 2004/0024583, entitled *Natural-Language Processing Using a Large Corpus*, filed by Robert J. Freeman on March 20, 2001 ("Freeman Application"). The Examiner has combined this new reference with the Zimmerman Patent as the bases for the current rejections. While the Examiner did not specifically address the Applicants' previous arguments supporting allowance, in light of the Examiner's reliance on a new reference and new bases for non-final rejections, Applicants conclude that their Nov. 12<sup>th</sup> Amendment and Remarks were persuasive and have overcome all previously outstanding rejections. Applicants now turn their attention to the new rejections in the present Office Action.

***Rejections Under 35 U.S.C. § 103***

Claims 1-11, 13, 16-24, 26 and 29-48 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over the Freeman Application in view of the Zimmerman Patent. Applicants respectfully traverse these rejections and request reconsideration.

***There is a Lack of Motivation to Combine the Freeman Application and Zimmerman Patent***

At the very least, there is no teaching, suggestion or motivation to combine the Freeman Application and Zimmerman Patents to render claims 1-11, 13, 16-24, 26 and 29-48 unpatentable. Obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either explicitly or implicitly in the references themselves or in the knowledge generally available to one of ordinary skill in the art. MPEP 2143.01. The Examiner has provided no permissible indication of any teaching, suggestion or motivation to combine these references.

The support for a motivation to combine that the Examiner does provide is flawed. The Examiner indicates that the motivation to combine is to "provide an efficient method for statistical analysis of a collection of topics that are related to both data corpus and the domain interest of the data analysis." Present Office Action at 4. The Examiner contends that a person of ordinary skill in the art would combine the "teachings of Freeman with the teachings of Zimmerman, wherein the subjects of the documents in the system provided therein (Freeman's figs 2 and sections 60 and 72),

would incorporate the use of topics of word combination, in the same conventional manner as described by Zimmerman (col. 8, lines 17-61)." *Id.*

A statement by an examiner that modifications of the prior art, such as that provided by the Examiner, to establish the claimed invention is not sufficient to establish a prima facie case of obviousness without some objective reason to combine the teachings of the reference. MPEP 2143.01 *citing Ex parte Levengood*, 28 USPQ2d 1300 (Bd. Pat. App. & Inter. 1993); *In re Kotzab*, 217 F.3d 1365, 1371, 55 USPQ2d 1313, 1318 (Fed. Cir. 2000).

To the extent that the Examiner attempts to provide an objective reason to support a motivation to combine, that reasoning is flawed on two levels. First, the Examiner has impermissibly defined the domain of interest, such that far too many references could conceivably be included within the domain of interest. Specifically, the Examiner has indicated that data and data analysis are the domain of interest, Present Office Action at 4, which covers an extraordinarily large range of subject matters and potential references. Defining such a large domain of interest serves notice that the Examiner is reaching to combine references that otherwise would not be permissibly combined, and is inconsistent with the notion that one skilled in the arts would reasonably combine references scattered across such a large domain of interest. For example, in the particular case, the Freeman Application focuses on natural language processing and the Zimmerman Patent is focused on information retrieval systems. One would be hard-pressed to find an individual skilled in both of these very different arts.

Second, respectfully, contrary to the Examiner's assertion above, there is no need for the systems in the Freeman Application to incorporate the methods related to

retrieving documents using topics (or topic generation) from the Zimmerman Patent.

The Freeman Application has no use for topics, so there would not be a motivation to use the Zimmerman Patent for the "use of topics of word combination," as suggested by the Examiner. Furthermore, the statistical measures used in the Zimmerman Patent amount to little more than counting how many times a word occurs and then listing the words in order of frequency. Zimmerman Patent at Col. 13, lines 64-67. One skilled in the art would not be motivated to turn to the Zimmerman Patent to incorporate such a simple statistical measure.

For at least these reasons there is no motivation to combine the Freeman Application and the Zimmerman Patent, therefore they can not be used as a basis of rejection. Reconsideration and allowance is respectfully requested of claims.

***The Cited References Fail to Serve as a Valid Basis for Rejection***

Assuming *arguendo* that the references could be combined, the combined references still fail to serve as a basis for rejection.

The Freeman Application discloses "[a] computer-parsing system based upon using vectors (lists) to represent natural language elements, providing a robust, distributed way to score grammaticality of an input string by using as a source material a large corpus of natural-language text." Freeman Application Abstract. "The system seeks to interpret natural language text for some Natural Language Processing applications, such as a speech recognition engine, a parser, a grammar checker, a search engine and so forth." *Id.* at ¶ 75. In short, to provide such an interpretation the Freeman Application discloses a means for scoring the grammatical correctness of a natural

language text string input. Unlike the Freeman Application, the Present Application focuses on the identification of topics within a data corpus, not natural language processing. Present Application Abstract.

The Zimmerman patent discloses a document retrieval system that contains a database that relates document word-pair patterns to topics. Zimmerman Patent Abstract. The system is initialized through training using a small initial database that has been manually indexed such that each document in the training database is manually assigned to one or more pre-existing categories or topics. *Id.* at col. 7, lines 1-4. Initially, documents are assigned to pre-existing topics. *Id.* at col. 4, lines 64-65. New documents added to the system are assigned topics by comparison of word pairing similarities in the new documents to the existing documents in the system. Topics are not word pairings that occur within a document, rather word pairings within a document are used to associate a document with a topic that is not the same as the word pairing. *Id.* at col. 14, lines 56-63. In fact within Zimmerman, many topics are identified as one word topics. *Id.* at col. 15, lines 20-25, *see also*, *Id.* at col. 4, line 52, *Id.* at Fig. 2. The Zimmerman Patent does not generate topics, nor is a word pairing in a document the same as a topic within the document. Unlike the Zimmerman Patent, the Present Application does generate topics and those topics are word combinations appearing in the documents that are indicated to have the particular topics.

***Claim 1***

***The cited references do not disclose or suggest the element of claim 1 of determining a segment-level actual usage value for one or more word combinations, wherein a word combination includes two or more substantially contiguous words, wherein two words are substantially contiguous if they are separated by zero words or words selected from a predetermined list of words***

With respect to claim 1, the Examiner alleges that the Freeman Application teaches the claimed feature of determining a segment-level usage value for one or more word combinations, which occurs within claim 1 of the present application. In support of this view, the Examiner refers to paragraphs 19, 48, and 60-72 in the Freeman Application. Second Office Action at 3. Applicants respectfully disagree with this view.

Applicants demonstrate below that none of these paragraphs teach the determining a segment-level usage value for one or more word combinations element. For illustrative purposes to show that the Freeman Application does not teach this element, one example of determining a segment-level usage value is provided. A segment-level usage value for a word combination, when a segment is defined as a paragraph, is the number of unique paragraphs in a data corpus in which the word-combination is found. See Present Application at ¶ 29.

Paragraph 19 is contained in the Background section of the Freeman Application, and reads as follows:

Statistical methods also keep a list of all possible relationships between all possible classes, but in [sic] they simplify model building by considering, in general, more simpler relationships, and use statistics to summarize regularities and optimize predictions. E.g., in +Spain => posit a statistical variable PREP where ("in", "on") are members of PREP and "Spain" follows PREP with probability  $P(\text{PREP}|\text{"Spain"})$ . The analysis is as with rules, but now each combination (branch of the tree) has a probability. Freeman Application at ¶ 19.

Paragraph 19 does not teach determining a segment-level usage value for one or more word combinations. Rather, paragraph 19 simply describes an existing statistical method for showing relationships between possible classes of words. Paragraph 19 deals with classes (e.g., prepositions, articles, nouns) and not words. Probabilities related to relationships of classes of words do not disclose or suggest how to determine a segment-

level usage value for a word. Furthermore, determining a segment-level usage value requires counting occurrences of words on a segment level, where a segment within a data corpus of many documents can be documents, paragraphs, or titles, etc. The explanation provided in paragraph 19 provides no teaching or suggestion related to looking at word counts on a segment-by-segment basis. And, in fact, such an approach would be inconsistent with the general thesis of the Freeman Application, which requires that a large data corpus is necessary to meaningfully evaluate the grammatical correctness of an input string and determine reasonably accurate results.

Paragraph 48 is contained within the Summary of Invention section of the Freeman Application, and reads as follows (note that the paragraph has been divided into sections to facilitate understanding the paragraph):

Yet in addition, in accordance with a preferred embodiment thereof,

this invention provides a computer system, using a provided corpus of linear natural-language elements of natural language test string data in a subject language and an input string of natural language elements in the subject language, comprising, in combination:

for a first adjoining pair, comprising a first pair element and a second pair element, of such natural language elements of such input string,

finding, from such string data from such corpus, a first listing of each such element syntactically equivalent to such first pair element and a second listing of each such element syntactically matching each such second pair element;

and from matching each such first-listing element with each such second-listing element, making a matched-pairs third listing by finding which matched pairs of said matching are found in such string data from such corpus;

wherein at least one of said adjoining pair comprises at least a pair of natural language elements; and,

further wherein at least one of such first pair element and such second pair element comprises at least a pair of words.

And it provides such a system wherein each such pair element comprises at least one word; and, further, wherein each such pair element comprises at least two words. Freeman Application at ¶ 48.



As with Paragraph 19, Paragraph 48 does not teach or suggest determining a segment-level usage value for one or more word combinations. Paragraph 48 discloses a method for determining a listing of pairs of words in a data corpus that are related to a pair of words in an input data string. The listing is created by matching words in the input string to syntactically related words in the data corpus. The description in paragraph 48 is silent on the generation of a usage value. In sum, in Paragraph 48 there is no counting of occurrences of words, no usage value determined, nor is the analysis done on a segment basis.

Paragraphs 60-72 also do not teach or suggest determining a segment-level usage value for one or more word combinations element. Paragraphs 60-70 simply refer to various tables and have no relationship to the above element. The tables show various combinations of words, but do not demonstrate the generation of a segment-level usage value.

Paragraphs 71 and 72 describe the improvements of the Freeman invention over the prior art. In particular, Freeman highlights that "recombination of vector elements to create new vectors associated with novel combinations gives this system a new power to describe idiosyncracies of linguistic association." Freeman Application at ¶ 72. The present application is not directed to "idiosyncracies of linguistic association," but, in fact, pertains to quite the opposite - generating meaningful topics that are not idiosyncratic, but useful and relatively common word combinations for identifying topics to associate with documents.

Furthermore, paragraph 72, as well as the other paragraphs cited by the Examiner, draws attention to the fact that the Freeman Application is very

focused on what the Present Application refers to as STOP Words, such as prepositions and articles. The consideration of prepositions and articles is of fundamental importance throughout the Freeman Application (and in natural language processing generally). The Present Application treats STOP words in an opposite way - they are eliminated from consideration as topics.

A review of the remainder of the Freeman Application demonstrates that nowhere in the Freeman Application does the amended determining a segment-level actual usage value element of claim 1.

***The cited references do not teach or suggest the element of claim 1 of computing a segment-level expected usage value for each of the one or more word combinations.***

The Examiner alleges that the Freeman Application teaches the element of computing a segment-level expected usage value for each of the one or more word combinations, which occurs within claim 1 of the present application. In support of this view, the Examiner refers to paragraph 72 in the Freeman Application. Second Office Action at 3. Applicants respectfully disagree with this view.

As discussed above, paragraph 72 describes the improvements of the Freeman Invention over the prior art. Nothing in paragraph 72 or the rest of the Freeman Application discloses or suggests that a segment level expected usage value is determined for word combinations that are substantially contiguous.

**The cited references do not teach or suggest the element of claim 1 of designating a word combination as a topic if the segment-level actual usage value of the word combination is greater than the segment-level expected usage value of the word combination**

The Examiner alleges that the Freeman Application teaches the claimed feature of designating a word combination if the segment-level actual usage value of the word combination is greater than the segment-level expected usage value of the word combination, which occurs within claim 1 of the present application. In support of this view, the Examiner refers to paragraphs 121-123 in the Freeman Application. Second Office Action at 3. Applicants respectfully disagree with this view.

Paragraphs 121-123 explain how a group of word pairings can be determined to be more grammatical than another group of word pairings. Freeman Application at ¶121. Such an approach does not teach or suggest how to designate a word combination as a topic. In fact, the approach for assessing word pairings in the Freeman Application teaches away from that in the Present Application in that the Freeman Application discloses comparing word pairings against different word pairings. In contrast, in the Present Application actual usage of a word combination is compared against the expected usage value of the same word combination. There is no comparison between different word combinations.

The Examiner acknowledges that the Freeman Application does not teach the identification of topics. Present Office Action at 4. Because the Freeman Application makes absolutely no mention of how to identify topics, the Examiner attempts to remedy this by asserting that the Zimmerman Patent teaches analysis and categorization of the words or word combinations of the topics. *Id.* Applicants respectfully disagree, the Zimmerman Patent does not teach identification of topics. Rather the Zimmerman Patent starts with a predetermined list of topics and assigns documents to those topics based on either word pairings or comparison of documents. This is exactly opposite to the invention of the Present Application. The Present Applications does not start with a

predetermined list of topics, rather the Present Application identifies topics and assigns those topics to documents that they occur in.

There are numerous other fundamental differences between the Zimmerman Patent and the Present Application. For example, in the Present Application topics include word combinations that are two or more words that occur in the documents they are assigned to. In the Zimmerman Patent topics are most commonly one word topics, and the topics do not necessarily occur in the documents for which they are assigned. The word pairings identified in the Zimmerman Patent are not the same as topics. This contrasts with the approach in the Present Application in which the word combinations that meet the statistical tests become a topic.

Moreover, at the very least, the Zimmerman Patent does not teach, disclose or suggest the element of computing a segment-level expected usage value for a word combination element of claim 1 of the present application, nor does it teach, disclose or suggest the element of designating a word combination as a topic if the segment level usage value of the word combination is greater than the segment level usage value of the word combination element of claim 1.

Numerous reasons have been provided above that show that the Freeman Application, the Zimmerman Patent or the combination of the two do not teach, disclose or suggest the elements of claim 1 of the present application. For any one of those reasons, independent claim 1 is patentable over the Freeman Patent Application and Zimmerman Patent. Reconsideration and allowance of this claim is respectfully requested.

**Claims 2-11 and 13**

Claims 2-5, 8-11 and 13 depend upon claim 1. Because each dependent claim incorporates all of the elements of the independent claim from which it depends, as well as additional features, the above arguments made with respect to claim 1, apply *a fortiori* to the dependent claims. For at least this reason, claims 2-11 and 13 are also patentable over the Freeman Patent Application, the Zimmerman Patent and their combination. Reconsideration and allowance of these claims is respectfully requested.

**Claim 16**

The Examiner has applied the same basis for rejection to claim 16, as for claim 1. For at least any one of those reasons stated above with respect to claim 1, independent claim 16 is patentable over the Freeman Patent Application and Zimmerman Patent. Reconsideration and allowance of this claim is respectfully requested.

**Claims 17-19 and 22-26**

Claims 17-19 and 22-26 depend upon claim 16. Because each dependent claim incorporates all of the elements of the independent claim from which it depends, as well as additional features, the above arguments made with respect to claim 16, apply *a fortiori* to the dependent claims. For at least this reason, claims 17-19 and 22-26 are also patentable over the Freeman Patent Application, the Zimmerman Patent and their combination. Reconsideration and allowance of this claim is respectfully requested.

**Claim 29**

Claim 29 includes the element of identifying topics by a method as in claim 1, 11 and 12. As shown above each of claims 1, 11 and 12 are patentable over the Freeman Application, the Zimmerman Patent and their combination. As a result, for at least this reason, claim 29 is also patentable over the Freeman Patent Application and Zimmerman Patent. Reconsideration and allowance of this claim is respectfully requested.

**Claims 30-38**

Claims 30-38 depend upon claim 29. Because each dependent claim incorporates all of the elements of the independent claim from which it depends, as well as additional features, the above arguments made with respect to claim 29, apply *a fortiori* to the dependent claims. For at least this reason, thus, claims 30-38 are also patentable over the Freeman Patent Application, the Zimmerman Patent and their combination. Reconsideration and allowance of this claim is respectfully requested.

**Claim 39**

Claim 39 includes the element of that identifies topics by a method as in claim 1, 11 and 12. As shown above each of claims 1, 11 and 12 are patentable over the Freeman Application, the Zimmerman Patent and their combination. As a result, for at least this reason, claim 39 is also patentable over the Freeman Patent Application and Zimmerman Patent. Reconsideration and allowance of this claim is respectfully requested.

#### **Claims 40-48**

Claims 40-48 depend upon claim 39. Because each dependent claim incorporates all of the elements of the independent claim from which it depends, as well as additional features, the above arguments made with respect to claim 39, apply *a fortiori* to the dependent claims. For at least this reason, thus, claims 40-48 are also patentable over the Freeman Patent Application, the Zimmerman Patent and their combination. Reconsideration and allowance of this claim is respectfully requested.

Related to all dependent claims, the absence of a direct response to the Examiner's remarks regarding each of the dependent claims should not be interpreted as Applicants' agreement or acquiescence with respect to any of the Examiner's basis for rejection of a particular dependent claim.

#### **New Claims 49 and 50**

New Claim 49 is based on claim 1 with an additional element that states that the segment-level expected usage value for a word combination is based on the frequency counts of words within the word combination within the data corpus or a portion of the data corpus. Claim 50 is based on claim 29, and refers to display of topics generated with the method of claim 49. For the reasons discussed above with respect to claims 1 and 29, Applicants respectfully state that these claims are allowable over the Freeman Application and Zimmerman Patent.

***Conclusion***

All of the stated grounds of objection and rejection have been properly traversed, accommodated, or rendered moot. Applicants therefore respectfully request that the Examiner reconsider all presently outstanding objections and rejections and that they be withdrawn. Applicants believe that a full and complete reply has been made to the outstanding Office Action and, as such, the present application is in condition for allowance. If the Examiner believes, for any reason, that personal communication will expedite prosecution of this application, the Examiner is invited to telephone the undersigned at the number provided.

Prompt and favorable consideration of this Amendment and Reply is respectfully requested.

Respectfully submitted,

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.



Michael D. Specht  
Attorney for Applicants  
Registration No. 54,463

Date: 8/9/05

1100 New York Avenue, N.W.  
Washington, D.C. 20005-3934  
(202) 371-2600